

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-212854

(43)Date of publication of application : 06.08.1999

(51)Int.Cl.

G06F 12/00

G06F 13/00

G06F 17/30

G06F 17/40

(21)Application number : 10-011404

(71)Applicant : AI SOFT KK

(22)Date of filing : 23.01.1998

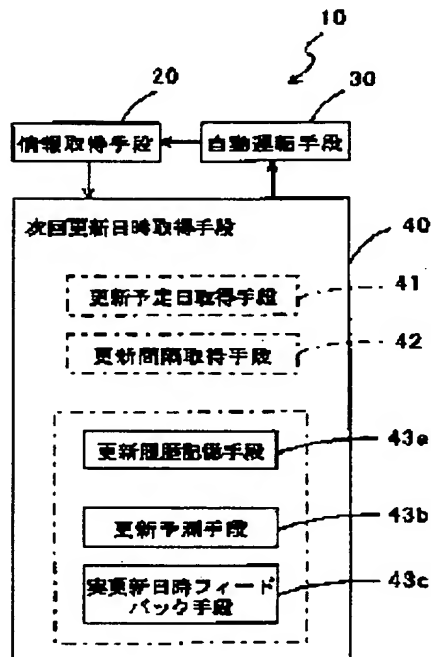
(72)Inventor : TOGASHI HIRONORI
SATO AKIHIRO
MOCHIDA AKIRA

(54) INFORMATION ACQUIRING DEVICE AND METHOD, AND MEDIUM STORING INFORMATION ACQUIRING PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To eliminate the useless accesses and also not to miss the latest information as much as possible by analyzing the acquired information to estimate the next updating date.

SOLUTION: An information acquiring means 20 of an information acquiring device 10 actively accesses an information originating source via a wide band network and a communication circuit and acquires the information from the information originating source when the means 20 is connected to the source. An automatic operation means 30 performs an automatic operation to secure the connection to a prescribed information originating source on a set date via the means 20. A next updating date acquiring means 40 analyzes the information acquired by the means 20 to acquire the next updating date and also sets the checking date of the means 30 to acquire the information after the next updating date. In such cases, the wide band network corresponds to an internet, etc., and the information originating source corresponds to a so-called server.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-212854

(43)公開日 平成11年(1999) 8月6日

(51)Int.Cl. ⁵	識別記号	F I
G 0 6 F 12/00	5 4 6	G 0 6 F 12/00 5 4 6 P
13/00	3 5 5	13/00 3 5 5
17/30		15/40 3 1 0 F
17/40		3 1 0 C
		15/401 3 4 0 A

審査請求 未請求 請求項の数13 O L (全 13 頁) 最終頁に続く

(21)出願番号 特願平10-11404

(22)出願日 平成10年(1998) 1月23日

(71)出願人 594067221

エー・アイ ソフト株式会社

長野県諏訪市大和三丁目3番5号

(72)発明者 富樫 宏規

長野県松本市中央二丁目1番27号 エー・

アイ ソフト株式会社内

(72)発明者 佐藤 昭弘

長野県松本市中央二丁目1番27号 エー・

アイ ソフト株式会社内

(72)発明者 持田 晃

長野県松本市中央二丁目1番27号 エー・

アイ ソフト株式会社内

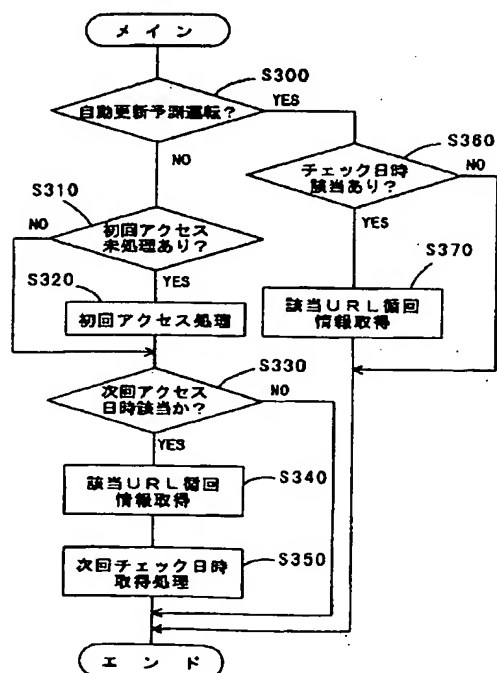
(74)代理人 弁理士 鈴木 喜三郎 (外2名)

(54)【発明の名称】 情報取得装置、情報取得方法および情報取得プログラムを記録した媒体

(57)【要約】

【課題】 情報発信源は不特定多数の利用者に対して更新情報を発信することは現実的に不可能であり、アクセスする側において情報が更新されたか否かを判断しなければならなかった。

【解決手段】 従来であれば、勘に頼って定期的な自動運転を行ない、当てずっぽう式の手法で情報を取得するようにしていたが、得られた情報の中に次の更新予定日であるとか、更新間隔などがないか判断するとともに、これらが無い場合には過去の更新履歴に基づいて次の更新を予測するようにしたため、いたずらにアクセス回数を増やすことなく、可能な限り最新の情報を維持することが可能となる。



【特許請求の範囲】

【請求項 1】 広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得装置であって、上記通信回線を介して上記情報発信源と接続されたときに当該情報発信源が発信している情報を取得する情報取得手段と、

設定したチェック日時に上記情報取得手段を所定の情報発信源に接続させるように自動運転する自動運転手段と、

上記情報取得手段によって取得した上記情報を解析することによって次回更新日時を得るとともに当該次回更新日時後に上記情報を取得するように上記自動運転手段のチェック日時を設定する次回更新日時取得手段とを具備することを特徴とする情報取得装置。

【請求項 2】 上記請求項 1 に記載の情報取得装置において、上記次回更新日時取得手段は、取得した上記情報を解析して次回の更新予定日があれば当該更新予定日を次回更新日時と判断する更新予定日取得手段を有することを特徴とする情報取得装置。

【請求項 3】 上記請求項 1 または請求項 2 のいずれかに記載の情報取得装置において、上記次回更新日時取得手段は、

取得した上記情報を解析して直近の更新日時と更新間隔の情報を取得して上記直近の更新日時から当該更新間隔後の日時を次回更新日時と判断する更新間隔取得手段を有することを特徴とする情報取得装置。

【請求項 4】 上記請求項 1 ～請求項 3 のいずれかに記載の情報取得装置において、上記次回更新日時取得手段は、

取得した上記情報の内容から得られる過去の更新日時情報を記憶する更新履歴記憶手段と、この更新履歴記憶手段にて記憶されている更新日時情報の履歴に基づいて次回更新日時を予測する更新予測手段と、

この予測結果に基づく次回のチェック日時に得られる上記情報に基づいて実更新日時を得て上記更新予測手段による次回更新日時の予測にフィードバックをかける実更新日時フィードバック手段とを有することを特徴とする情報取得装置。

【請求項 5】 上記請求項 1 ～請求項 4 のいずれかに記載の情報取得装置において、上記更新履歴記憶手段は、直近の更新日時とともに過去の複数回の更新履歴を記憶し、

上記更新予測手段は、上記更新履歴から更新間隔を取得しつつ直近の更新日時に同更新間隔を加えたものを次回更新日時と予測することを特徴とする情報取得装置。

【請求項 6】 上記請求項 5 に記載の情報取得装置において、上記更新予測手段は、上記更新履歴における間隔の平均を更新間隔とすることを特徴とする情報取得装

置。

【請求項 7】 上記請求項 5 に記載の情報取得装置において、上記更新予測手段は、上記更新履歴における間隔の最小値を更新間隔とすることを特徴とする情報取得装置。

【請求項 8】 上記請求項 5 に記載の情報取得装置において、上記更新予測手段は、最初の情報取得時に直近の更新日時と当該情報取得日時との間隔を取得して更新間隔と判断することを特徴とする情報取得装置。

10 【請求項 9】 上記請求項 1 ～請求項 8 のいずれかに記載の情報取得装置において、上記情報発信源は W E B サイトであり、上記情報は H T M L 文書であることを特徴とする情報取得装置。

【請求項 10】 上記請求項 1 ～請求項 8 のいずれかに記載の情報取得装置において、上記情報発信源は、当該情報取得端末で実行するプログラムファイルの供給サーバであり、上記情報には、当該プログラムファイルが含まれ、実行用プログラムファイルの最新情報を取得することを特徴とする情報取得装置。

20 【請求項 11】 上記請求項 1 ～請求項 10 のいずれかに記載の情報取得装置において、上記広域ネットワークが、インターネットであることを特徴とする情報取得装置。

【請求項 12】 広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得方法であって、設定したチェック日時に自動運転して、上記通信回線を介して上記情報発信源と接続し、当該情報発信源が発信している情報を取得するとともに、この取得した情報を解析することによって次回更新日時を得て、この更新日時後に上記情報を取得するように次回のチェック日時を設定することを特徴とする情報取得方法。

【請求項 13】 広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得プログラムを記録した媒体であって、

設定したチェック日時に自動運転して、上記通信回線を介して上記情報発信源と接続し、当該情報発信源が発信している情報を取得するとともに、この取得した情報を解析することによって次回更新日時を得て、この更新日時後に上記情報を取得するように次回のチェック日時を設定することを特徴とする情報取得プログラムを記録した媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、情報取得装置、情報取得方法および情報取得プログラムを記録した媒体に関し、特に、広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得装置、情報取得方法およ

び情報取得プログラムを記録した媒体に関する。

【0002】

【従来の技術】近年、インターネットのような広域ネットワークにおいては、極めて多数の情報発信源が、おのおの情報を発信しつつある。このような情報発信源の特徴として、一定期間毎に発信情報を更新していることが多い。

【0003】この場合、情報発信源は不特定多数の利用者に対して更新情報を発信することは現実的に不可能であり、アクセスする側において情報が更新されたか否かを判断しなければならない。

【0004】

【発明が解決しようとする課題】上述した従来の情報取得装置においては、次のような課題があった。

【0005】アクセスする側では更新の有無が分からないため、自ら能動的にアクセスして更新の有無をチェックしなければならない。しかしながら、頻繁にチェックすることにすれば、更新がない限りは何度も無駄にアクセスすることが生じるし、十分な期間をおいてチェックすることにすれば最新の情報を取得できなくなってしまう。

【0006】本発明は、上記課題にかんがみてなされたもので、効率よく最新の情報を得ることが可能な情報取得装置、情報取得方法および情報取得プログラムを記録した媒体の提供を目的とする。

【0007】

【課題を解決するための手段】上記目的を達成するため、請求項1にかかる発明は、広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得装置であって、上記通信回線を介して上記情報発信源と接続されたときに当該情報発信源が発信している情報を取得する情報取得手段と、設定したチェック日時に上記情報取得手段を所定の情報発信源に接続させるように自動運転する自動運転手段と、上記情報取得手段によって取得した上記情報を解析することによって次回更新日時を得るとともに当該次回更新日時後に上記情報を取得するように上記自動運転手段のチェック日時を設定する次回更新日時取得手段とを具備する構成としてある。

【0008】上記のように構成した請求項1にかかる発明においては、予め設定したチェック日時に自動運転手段が情報取得手段を所定の情報発信源に接続させるように自動運転すると、当該情報取得手段はその情報発信源が発信している情報を取得する。一方、次回更新日時取得手段は、上記情報取得手段によって取得した上記情報を解析することにより、同情報発信源における次回更新日時を得て、当該次回更新日時後に上記情報を取得するように上記自動運転手段のチェック日時を設定する。従って、このチェック日時となると上述したような自動運転によって再度情報を取得していくことになる。

【0009】すなわち、アクセスして得られた情報を解析することにより、次回更新日時を自動的に予測し、その更新日時を目途にしたチェック日時にアクセスして情報を得る。この場合、実際にアクセスするのは予測された次回更新日時よりも後であるべきであるから、次回更新日時と次回チェック日時というのは異なるはずである。従って、厳密に言えば1秒でも後でなければならないということになるが、必ずしもこのような厳密な処理が必要であるとは言えない場合も多い。あくまでも予測という性格があるし、発信者の側が時間まで厳密に管理していないことも多いからである。従って、予測された次回更新タイミングは日にち単位の予定日としてとらえ、時間はその日の正午と把握したり、夜中の12時と考えることも可能である。

【0010】次回更新日時取得手段が実現するのは情報源における次回更新日時を取得することであり、これは情報内容によってさまざまなアプローチを取ることが可能である。その一例として、請求項2にかかる発明は、請求項1に記載の情報取得装置において、上記次回更新日時取得手段は、取得した上記情報を解析して次回の更新予定日があれば当該更新予定日を次回更新日時と判断する更新予定日取得手段を有する構成としてある。

【0011】上記のように構成した請求項2にかかる発明においては、取得した上記情報に基づいて次回更新日時取得手段が解析を行なう。この解析は同情報内に次回の更新予定日があるか否かを判定するものであり、あれば当該更新予定日を次回更新日時と判断する。すなわち、情報発信源が予定している更新日を公表している場合にはこれを利用している。

【0012】また、別のアプローチの一例として、請求項3にかかる発明は、請求項1または請求項2のいずれかに記載の情報取得装置において、上記次回更新日時取得手段は、取得した上記情報を解析して直近の更新日時と更新間隔の情報を取得して上記直近の更新日時から当該更新間隔後の日時を次回更新日時と判断する更新間隔取得手段を有する構成としてある。

【0013】上記のように構成した請求項3にかかる発明においても、取得した上記情報に基づいて次回更新日時取得手段が解析を行なうが、ここでは同情報内に更新間隔と直近の更新日時があるか否かを判定する。そして、これらがある場合には直近の更新日時から当該更新間隔後の日時を次回更新日時と判断する。すなわち、情報源が更新間隔を予定している場合にはこれを利用している。

【0014】さらに、別のアプローチとして、請求項4にかかる発明は、請求項1～請求項3のいずれかに記載の情報取得装置において、上記次回更新日時取得手段は、取得した上記情報の内容から得られる過去の更新日時情報を記憶する更新履歴記憶手段と、この更新履歴記憶手段にて記憶されている更新日時情報の履歴に基づい

て次回更新日時を予測する更新予測手段と、この予測結果に基づく次のチェック日時に得られる上記情報に基づいて実更新日時を得て上記更新予測手段による次回更新日時の予測にフィードバックをかける実更新日時フィードバック手段とを有する構成としてある。

【0015】上記のように構成した請求項4にかかる発明においても、取得した上記情報に基づいて次回更新日時取得手段が解析を行なう。この場合、次回更新日時取得手段は、更新履歴記憶手段と、更新予測手段と、実更新日時フィードバック手段とを備えており、更新履歴記憶手段が取得した上記情報の内容から得られる過去の更新日時情報を記憶するので、更新予測手段は記憶されている更新日時情報の履歴に基づいて次回更新日時を予測する。一方、このように予測された更新日時に対応して次回チェック日時にアクセスされたときには、必ずしも予測どおりとはいえないことがありえる。このため、実更新日時フィードバック手段は、次のチェック日時に新たに得られる上記情報に基づいて実更新日時を得て、上記更新予測手段による次回更新日時の予測にフィードバックをかける。

【0016】このフィードバックの手法は更新予測手段による予測手法に応じて様々な態様が可能である。この更新予測手段における更新予測の一例として、請求項5にかかる発明は、請求項1～請求項4のいずれかに記載の情報取得装置において、上記更新履歴記憶手段は、直近の更新日時とともに過去の複数回の更新履歴を記憶し、上記更新予測手段は、上記更新履歴から更新間隔を取得しつつ直近の更新日時に同更新間隔を加えたものを次回更新日時と予測する構成としてある。

【0017】上記のように構成した請求項5にかかる発明においては、更新予測手段が過去の複数の更新履歴から次回更新日時を予測するにあたって、更新履歴記憶手段が記憶するそれぞれの更新履歴の間隔から更新間隔を演算し、直近の更新履歴に更新間隔を加算している。

【0018】このようにして更新日時を予測したとしても必ずしも予定どおりにいくとは限らない。従って、この更新間隔が過ぎた時点でアクセスしても更新されていない場合は、実更新日時フィードバック手段がこれまでの更新間隔の変動を解析して更新間隔の変動度合いを考慮した次回更新日時を得ればよい。そして、このような変動度合いを考慮に入れた新たな手法で次回更新日時を決定させる。むろん、あまり複雑とさせることなく、単純に次回更新日時を翌日にさせるといったものであっても良い。

【0019】一方、既に更新されていた場合にはその実更新日時に基づいて更新間隔をより短くすればよい。

【0020】また、上述した更新間隔の求め方についても各種の手法を採用可能であり、その一例として、請求項6にかかる発明は、請求項5に記載の情報取得装置において、上記更新予測手段は、上記更新履歴における間

隔の平均を更新間隔とする構成としてある。すなわち、更新履歴の間隔の平均を更新間隔としている。

【0021】また、他の一例として、請求項7にかかる発明は、請求項5に記載の情報取得装置において、上記更新予測手段は、上記更新履歴における間隔の最小値を更新間隔とする構成としてある。すなわち、更新履歴の間隔の最小値を更新間隔としている。

【0022】以上のようにして更新間隔を利用するとした場合、複数の更新履歴が無いときには更新間隔を得られないこともあり得る。すなわち、最初の情報取得時には一つしか直近の更新日時だけしか得られないことがある。このような場合に好適な一例として、請求項8にかかる発明は、請求項5に記載の情報取得装置において、上記更新予測手段は、最初の情報取得時に直近の更新日時と当該情報取得日時との間隔を取得して更新間隔と判断する構成としてある。

【0023】上記のように構成した請求項8にかかる発明においては、上記更新予測手段が最初の情報取得時に得た直近の更新日時と当該情報取得日時との間隔を取得し、当該間隔を更新間隔と判断する。

【0024】むろん、最初の情報取得時に過去の更新履歴が複数得られればこれを利用することも可能である。

【0025】一方、情報取得手段が広域ネットワークを介して情報を取得する情報発信源は様々であり、特定のものに限られる必要はない。しかしながら、近年においてはインターネットによる情報発信が盛んであり、広域ネットワークとしてインターネットを利用した情報取得にも適用可能であることはいうまでもない。このような場合に好適な一例として、請求項9にかかる発明は、請求項1～請求項8のいずれかに記載の情報取得装置において、上記情報発信源はWEBサイトであり、上記情報はHTML文書で構成してある。

【0026】上記のように構成した請求項9にかかる発明においては、情報取得手段は広域ネットワークを介して情報発信源であるWEBサイトにアクセスし、同WEBサイトが発信するHTML文書の情報を取得する。むろん、このHTML文書はドキュメントとしての文書情報であっても良いし、図形や写真データであっても良いし、プログラムファイルなどであっても良い。さらに、サーバスクリプトであってもよい。さらに、CGIスクリプトを実行させることによって返信されるHTML文書なども当然に含まれる。また、HTML文書を取得する際にはHTTPプロトコルが採用され、その過程でも情報のやりとりが行われる。この例で言えば、本来のHTML文書とともに拡張情報として最終更新日時が「LastModified: Thu, 01 Jan 1998 10:23:00 GMT」というようにして取得可能である。従って、このようにして取得される情報も広義の意味での上記情報に含まれることはいうまでもない。

【0027】これに対し、情報発信源はこのような一般

向けのHTML文書などに限られるものではなく、システムファイルのアップデートプログラムのサーバなどにも適用可能である。その一例として、請求項10にかかる発明は、請求項1～請求項8のいずれかに記載の情報取得装置において、上記情報発信源は、当該情報取得端末で実行するプログラムファイルの供給サーバであり、上記情報には、当該プログラムファイルが含まれ、実行用プログラムファイルの最新情報を取得する構成としてある。

【0028】上記のように構成した請求項10にかかる発明においては、情報発信源が情報取得端末で実行するプログラムファイルの供給サーバであり、いわゆる最新のアップデートやプログラムファイルを供給している。従って、情報取得端末の側では当該供給サーバから最新のプログラムファイルを得られるようにその更新状況を把握する必要がある。従って、上述したように次回更新日時を予測してアクセスし、更新されたプログラムファイルがあればこれを取得することになる。

【0029】そして、これらの場合においては広域ネットワークがインターネットである場合も多く、請求項11にかかる発明は、請求項1～請求項10のいずれかに記載の情報取得装置において、上記広域ネットワークが、インターネットとなっている。

【0030】自動運転手段の構成もさまざまなものを採用可能である。コンピュータであれば、情報取得プログラムと一体的に構成され、その自動運転機能として実現されるものであっても良い。また、コンピュータに対するスクリプトとして個々の情報取得プログラムの起動や操作をコントロールするものであっても良い。さらに、複数の情報取得端末が接続されるプロキシサーバなどにおいて、広域ネットワークに接続して必要な情報を取得し、取得した情報を蓄えておいて各情報取得端末に配信するものであってもよい。

【0031】このように、取得した情報に基づいて更新日時を予測して自動運転のチェック日時を設定する手法は必ずしも実体のある装置に限られる必要はなく、その方法としても機能することは容易に理解できる。このため、請求項12にかかる発明は、広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得方法であって、設定したチェック日時に自動運転して、上記通信回線を介して上記情報発信源と接続し、当該情報発信源が発信している情報を取得するとともに、この取得した情報を解析することによって次回更新日時を得て、この更新日時後に上記情報を取得するように次回のチェック日時を設定する構成としてある。

【0032】すなわち、必ずしも実体のある装置に限らず、その方法としても有効であることに相違はない。

【0033】ところで、このような情報取得装置は単独で存在する場合もあるし、ある機器に組み込まれた状態

で利用されることもあるなど、発明の思想としてはこれに限らず、各種の態様を含むものである。従って、ソフトウェアであったりハードウェアであったりするなど、適宜、変更可能である。

【0034】発明の思想の具現化例として情報取得装置のソフトウェアとなる場合には、かかるソフトウェアを記録した記録媒体上においても当然に存在し、利用されるといわざるをえない。

【0035】その一例として、請求項13にかかる発明は、広域ネットワークの通信回線を介して情報発信源に対して能動的にアクセスして同情報発信源が発信する情報を取得する情報取得プログラムを記録した媒体であって、設定したチェック日時に自動運転して、上記通信回線を介して上記情報発信源と接続し、当該情報発信源が発信している情報を取得するとともに、この取得した情報を解析することによって次回更新日時を得て、この更新日時後に上記情報を取得するように次回のチェック日時を設定する構成としてある。

【0036】むろん、その記録媒体は、磁気記録媒体であってもよいし光磁気記録媒体であってもよいし、今後開発されるいかなる記録媒体においても全く同様に考えることができる。また、一次複製品、二次複製品などの複製段階については全く問う余地無く同等である。その他、供給方法として通信回線を利用して行なう場合でも本発明が利用されていることにはかわりない。

【0037】さらに、一部がソフトウェアであって、一部がハードウェアで実現されている場合においても発明の思想において全く異なるものではなく、一部を記録媒体上に記憶しておいて必要に応じて適宜読み込まれるような形態のものとしてあってもよい。

【0038】

【発明の効果】以上説明したように本発明は、取得した情報を解析して次回更新日時を予測するようにしたため、無駄にアクセスする必要がなくなるとともに、できるだけ最新の情報を逃さないようにすることが可能な情報取得装置を提供することができる。

【0039】また、請求項2にかかる発明によれば、情報発信源が発信する更新予定を利用するため、情報発信源の信頼性がある限り、正確で最新の情報を無駄なく取得することができる。

【0040】さらに、請求項3にかかる発明によれば、情報発信源が発信する更新間隔に基づいて次回更新日時を推測するため、この場合も正確で最新の情報を無駄なく取得することができる。

【0041】さらに、請求項4にかかる発明によれば、過去の複数の更新履歴を記憶することにより、情報発信源が直接に更新間隔を発信していない場合でも、更新間隔を取得することができ、効率的な情報取得が可能となる。

【0042】さらに、請求項5にかかる発明によれば、

直近の更新日時に更新間隔を加えるというシンプルな手法で次回更新日時を予測することができる。

【0043】さらに、請求項6にかかる発明によれば、採用するのが更新間隔の平均であるので、アクセスのロスと最新情報の取得可能性とのバランスが安定するといえる。

【0044】さらに、請求項7にかかる発明によれば、採用する更新間隔が最小値であれば次回チェック日時に未更新である確率が高いが、最新の情報を得やすくなると言える。

【0045】さらに、請求項8にかかる発明によれば、最初の情報取得時であっても更新間隔を擬似的に作り出して実行することができる。

【0046】さらに、請求項9にかかる発明によれば、WEBサイトから効率よくHTML文書を取得することができる。

【0047】さらに、請求項10にかかる発明によれば、プログラムファイルの供給サーバから効率よく最新のプログラムファイルを取得することができる。

【0048】さらに、請求項11にかかる発明によれば、インターネットで様々な最新の情報を無駄なく取得することができる。

【0049】さらに、請求項12にかかる発明によれば、上記のような効果を奏する情報取得方法を提供でき、請求項13にかかる発明によれば、情報取得プログラムを記録した媒体を提供することができる。

【0050】

【発明の実施の形態】以下、図面にもとづいて本発明の実施形態を説明する。

【0051】図1は本発明の一実施形態にかかる情報取得装置をクレーム対応図により示しており、図2は本情報取得装置が広域ネットワークに接続される全体像を概略的に示している。

【0052】本情報取得装置10は、広域ネットワーク50と通信回線60を介して情報発信源70に対して能動的にアクセスして上記情報発信源70と接続されたときに当該情報発信源70が発信している情報を取得する情報取得手段20と、設定したチェック日時に上記情報取得手段20にて所定の情報発信源70に接続させるように自動運転する自動運転手段30と、上記情報取得手段20によって取得した上記情報を解析することによって次回更新日時を得るとともに当該次回更新日時後に上記情報を取得するように上記自動運転手段30のチェック日時を設定する次回更新日時取得手段40とを備えている。

【0053】広域ネットワーク50はインターネットなどが該当するし、情報発信源70はいわゆるサーバが該当する。すなわち、情報発信源70は所定のURLアドレスを割り当てられたサーバを備えており、当該URLアドレスに基づいて情報取得装置10が能動的にアクセ

スしてきた場合に必要な情報を取得させる。

【0054】情報の発信、取得の形態はさまざまであるが、情報発信源70がWEBサイトであるとする、発信する情報は概ねHTML文書であり、この中にはテキスト文書の他、画像ファイル、プログラムファイル、音声ファイルなど、各種の情報が含まれている。

【0055】一方、このような広域ネットワーク50に利用される通信回線60は電氣的なものや光利用のものが可能であるし、各種の媒体変換装置を介して接続されている。インターネットであればイーサネットケーブル61による直接接続という手法も可能であるし、あるいはプロバイダ62に対して公衆電話回線63を介してダイヤルアップ接続することも可能であり、この場合にはプロバイダ62や公衆電話回線63も通信回線60の一部に含まれることになる。むしろ社内LAN64を介して接続される場合にはこのような社内LAN64やルータ65なども通信回線60の一部に含まれることになる。

【0056】情報取得装置10自体は、図3に示すようなコンピュータ端末100で構成することができる。コンピュータ端末100は、CPU111やROM112やRAM113やI/O114などを備えるコンピュータ110とともに、モニタ120やキーボード130やマウス140といった主構成装置とともに、フロッピーディスクドライブ150や、ハードディスクドライブ160や、CD-ROMドライブ170といった外部記憶装置が含まれ、さらには、上述した通信回線60に接続するためのモデム180やLANアダプタ190などから構成されている。

【0057】コンピュータ端末100は上述したハードウェアユニットが後述するソフトウェアを実行して各種の処理を実現するものであるから、相互に一体となって情報取得装置10を構成していると言える。すなわち、上述したように用意されたハードウェア資源をソフトウェアによって最大限に利用することにより、不可分一体となって図1に示すような情報取得手段20や自動運転手段30や次回更新日時取得手段40を実現している。

【0058】図4および図5はこの情報取得装置10を実現するソフトウェアの概略フローチャートと設定用画面を示している。

【0059】図5の設定用画面200では、左半分程の領域を、巡回先URL指定エリア210と、各URL毎の更新履歴表示エリア220とが占めており、右半分程の領域を、自動更新予測指示領域230と、定期運転指示領域240とによって占めている。尚、定期運転指示領域240には、自動運転間隔を週単位や月単位で指示したり、一日のうちの運転時間を指示するための個別指示領域241～243が用意されている。例えば、週単位個別指示領域241では、週単位では任意の曜日を一つまたは複数指定できる。この例では、曜日を指定して

毎週運転するようにしているが、第1週であるとか第3週といった指定であっても良い。次に、月単位個別指示領域242では、月単位で任意の日付を四つまで指定できる。この例では四つまでとなっているが、31日分のチェックボックスを用意しておいて任意の数だけ指定できるようにしても良い。そして、日単位個別指示領域243では、毎日指定時刻になったときにアクセスするか、週単位や月単位でアクセスする場合の時間を指定する。

【0060】個別指示領域241～243については択一の選択が可能であり、また、自動更新予測指示領域230と定期運転指示領域240についても択一の選択が可能となっている。

【0061】このような設定用画面200での設定に基づいて図4に示す手順に従って情報取得の処理が実行される。最初に、ステップS300では択一的な指示が可能な自動更新予測運転か定期運転かを判断する。自動更新予測指示領域230と定期運転指示領域240とはボタンによって択一的な選択となっており、ここで自動更新予測指示領域230の側のボタンをセットしてあるときには、ステップS310へとすすむし、定期運転を指示しているときにはその指示内容に応じて該当するチェック日時であるか否かをステップS360にて判断し、該当する場合にはステップS370にて該当するURLのWEBサイトに対して能動的にアクセスし、情報を取得する。

【0062】ここで自動更新予測処理の実体について説明する。本自動更新予測処理の実質は、図6に示す初回アクセス処理と、図7に示す次回チェック日時取得処理とから構成される。

【0063】次回チェック日時取得処理は、最新の情報をできる限り少ないアクセスで確保しようとするものであり、WEBサイトから得られるHTML文書を解析して当該WEBサイトが次回に情報を更新する日時を予測する。この予測の手法は、概ね三種に分かれる。以下、このHTML文書を単に情報と呼ぶ。

【0064】第一の手法は、情報の中に次回更新日の開示が有るか否かを検討する手法である。WEBサイトの運営者が次回の更新予定日を情報中に開示することが多い。従って、同情報中に開示されていないか判断する。具体的には、文書データに絞り込み、「次回更新予定日」という文字データ、あるいはこれに類似する文字データがあるか否かを判断する。そして、見つかった場合にはそれに続く文字データとしての日付を取得する。日付は文字データとして表すときにいくつものフォーマットが考えられる。例えば、「1998年1月1日」も「平成10年1月1日」も「1998/1/1」も全て同じ日付であるため、整形する処理と合わせて日付を得る。そして、得られた日付を次回チェック日時として判断する。

【0065】なお、更新予定日が日にち単位でしか取得できない場合もあるが、その場合には正午であるとか、夜中の12時などに統一してしまう。また、以下においても、日にち単位の判断で行わなければならない場合もあるが、必ずしも厳密に考えることなく、大まかには日にち単位で処理すればよい。従って、内部的には時間まで設定しておくことにしたとしても、コンピュータ端末100がその時間に稼働しているとは限らないことまで考えれば、たとえ次回更新日時が午後7時であったとしても、日付が一致したならば午後3時に自動運転するということも有意義だからである。むしろ、原則どおりに、時間まで踏まえて制御し、午後7時前にコンピュータ端末100がオフにされたら、翌日に起動された時点でチェックするようにしても構わない。

【0066】この手法は図1において更新予定日取得手段41に該当し、図6の初回アクセス処理でのステップS405、S410に該当するとともに、図7の次回チェック日時取得処理でのステップS510、S515に該当する。

【0067】第二の手法は、情報の中に更新間隔の開示が有るか否かを検討する手法である。WEBサイトの運営者が更新間隔を定めていることが多く、そのような更新間隔の開示がないか判断する。具体的時には、情報の中に「毎週・・更新」、「毎月・・更新」、「毎・・日更新」といった文字データ、あるいはこれに類似する文字データがあるか否かを判断する。そして、見つかった場合にはその更新間隔を解析する。

【0068】また、WEBサイトでは少なくとも発信している情報の最終更新日時だけは開示していることが多い。従って、最終更新日時も次回更新予定日と同様に解析して取得し、最終更新日時に上述した更新間隔を加えた日を次回チェック日時として判断する。この手法は図1において更新間隔取得手段42に該当し、図6の初回アクセス処理でのステップS415、S420に該当するとともに、図7の次回チェック日時取得処理でのステップS520、S525に該当する。

【0069】ただし、この最終更新日時についてはWEBサイトからHTML文書を取得するときに拡張情報として同時に取得できる情報でもある。従って、かかる解析によらずにHTML文書を取得することによって得るようにしても構わない。以下、情報の中から取得することを前提とし、情報の中に含まれていない場合には補助的に同拡張情報から得ることにする。

【0070】第三の手法は、情報の中から更新履歴を読み取る手法である。このアプローチには更に二つの手法があり、一つ目は情報内に過去の複数の更新履歴が更新日時リストとして開示されている場合に、それらの間隔を見い出す手法であり、二つ目は最終の更新日時だけが開示されている場合にはそれを記憶して自らのデータベースとしておき、複数回情報を取得したときに間隔を見

出す手法である。

【0071】更新日時リストを得られたときには複数の更新間隔が得られることになるが、このときには最小の間隔をとるようにしてもよいし、平均的な間隔を演算によって求めるようにしても良い。最新の情報を求める希望を優先すれば最小の間隔を採用すべきであるものの、たまたまあるときに更新間隔が短かったという場合には無駄なアクセスが増えることになる。一方、平均の間隔を採用する場合には、無駄なアクセスを防止する傾向を高めるものの、あるときから更新間隔を短くしたという

ような場合には最新の情報を逃しやすくなる。従って、長期にわたる平均はむしろ弊害の方が大きいと考えられる。このため、最小の間隔としたり、直近の三回の更新間隔から平均を算出することとしたりする手法を適宜使い分けることにする。むろん、直近の三回がない場合には二回以下であっても構わない。

【0072】ところで、更新と同様の概念として掲載という概念もある。例えば、ニュースなどの情報があれば最終更新日時が分からなくても掲載日時リストがあればその間隔から更新間隔を実質的に把握できるということもある。もっとも、最終更新日時自体はHTML文書を取得するのと同時に得ることも可能である。

【0073】従って、以上のような更新日時リストや掲載日時リストがある場合には、その間隔の中の最小を更新間隔を求めるとともに、最終更新日時に加えて次のチェック日時とする。この手法は、図6においてステップS425～S440に該当する。

【0074】また、前者のものであれば初めて情報を取得したときにでも更新間隔を得られるが、後者のように最終更新日時しか得られないような場合には更新間隔は得られない。このため、今回チェック日時と得られた最終更新日時との間隔を擬似的にチェック間隔と置き換え、次回チェック日時は今回チェック日時にこのチェック間隔の半分を加える。この手法は、図6においてステップS445、S450に該当する。

【0075】初回アクセス処理の場合、ステップS455にて以上のようにして得られた次回チェック日時をそれぞれのURLごとに記憶する。また、過去の更新履歴が得られている場合にはそれも記憶するので、設定用画面200の更新履歴表示エリア220には各URL毎に更新履歴が表示される。例えば、過去の更新日時は枠内に四角のマーキングを行い、次回チェック日時には丸印のマーキングを行っている。

【0076】一方、次回以降のアクセスでは、過去三回の更新間隔の平均をとってチェック間隔とし、このチェック間隔を最終更新日時に加えて次回チェック日時とする。この手法をステップS530、S535に示している。

【0077】以上で、次回チェック日時を取得する手法について説明した。従って、図4に示すフローチャート

に戻れば、ステップS300にて自動更新予測運転と判断されると、まず、ステップS310にて以上のような初回アクセス処理が済んでいないかURLが登録されているかを判断する。これは設定用画面200では表示しなかったもののURLを設定するのに伴って用意されるフラグで判断できるようになっており、初期値は未処理となっている。従って、ステップS310にて各URLのフラグを判断し、未処理のものがあればステップS320にて、図6に示す初回アクセス処理を実行して次回チェック日時を取得する。

【0078】次に、ステップS330では各URLごとの次回チェック日時を参照し、本日に該当するものがあるかを判定する。次回チェック日時に該当するものがあればステップS340にて該当するURLのWEBサイトに対して能動的にアクセスし、情報を取得する。必要な情報を全て取得したら通信を終了し、ステップS350では図7に示す次回チェック日時取得処理を実行する。

【0079】次回チェック日時を取得する手法については上述したとおりであるが、この次回チェック日時取得処理を実行したのは予め予測した更新予定日であるから、実際に更新されていることもあれば実際には更新されていないこともあり得る。

【0080】従って、ステップS500にて情報を取得した後、ステップS505では情報が更新されているかを判断し、情報が更新されていれば上述したような手法で次のチェック日時を取得するとともに、ステップS545にて各URLの次回チェック日時にセットする。

【0081】これに対して、情報が更新されていない場合の処理が問題となる。この場合は、前回に設定した次回チェック日時が正しくないので、修正をかける。この修正の手法もいくつか有効であるが、本実施形態においては、最初にある程度の間隔を決定し、繰り返していくにつれて徐々に間隔を短くするという手法を採用している。その前提として、ステップS530では過去の更新間隔のうちの最小更新間隔を保存しておき、ステップS540ではこの最小更新間隔を連続未更新回数(+1)で割り、その商を今回のチェック日時に加えることにする。

【0082】また、初回アクセス処理においてステップS430、S440にて過去の更新日時リストなどから最小更新間隔を取得しておくとともに、過去の更新間隔を得られなかった場合にはステップS445にて最小更新間隔にチェック間隔を保存している。

【0083】このように、最小間隔を使用しながらも、徐々にその間隔を短くしながら次回チェック日時を修正するフィードバックをかけており、この意味ではWEBサイトの運営者が予定を開示していない場合、過去の更新履歴を記憶するとともに、その更新履歴に基づいて更

新予測を行い、さらには、実際の更新日時を得てフィードバックを行うことになる。すなわち、更新履歴記憶手段43a、更新予測手段43b、実更新日時フィードバック手段43cを実現しているといえる。

【0084】次に、上記構成からなる本実施形態の動作

$$\mu_i = r_h (1 - r_p) r_g / (1 + N) \times 2 \dots (1)$$

r_h : ヒット率 = 情報取得回数 / アクセス回数

r_p : ペナルティ率 = ペナルティ / 監視日数

(ペナルティ = 情報取得日時 - 最終更新日時)

r_g : 取得率 = 取得回数 / 更新回数

N : 情報喪失数

以下、図8のシミュレーションを実行することにする。あるWEBサイトでの実際の更新日は「1日」「5日」「11日」「22日」「27日」であったとし、「4日」から更新予測を行う自動設定パターンと、毎週月曜日と金曜日に定期運転する第一の定期パターンと、毎週火曜日に定期運転する第二の定期パターンとを対比する。なお、「4日」が月曜日である。

【0087】自動設定では、4日に情報を取得したところ、1日に最終更新日時があったことが分かった。従って、初回アクセス処理ではステップS445にて「チェック間隔 = 今回のチェック日時 - 最終更新日時」という計算を実行し、チェック間隔と最小更新間隔は「3日」とするとともに、次回チェック日時はステップS450の「次回チェック日時 = 今回のチェック日時 + チェック間隔 / 2」という計算を実行し、「6日」という結果を得る。なお、端数は切り上げとする。

【0088】この6日にアクセスすると、情報は「5日」に更新されていたため、これで二つの更新履歴を得られる。すると、ステップS530での「更新間隔 = 今回更新日時 - 前回更新日時」という計算を実行すると「4日」という結果が得られる。一方、チェック間隔は過去三回の更新間隔の平均であるが、一回分しかないのでそのまま利用すると「4日」となる。さらに、最小更新間隔も同様とする。そして、ステップS535にて「次回のチェック日時 = 最終更新日時 + チェック間隔」という計算を実行すると、「9日」となる。

【0089】しかし、9日にアクセスしても情報は更新されておらず、ステップS540では「次回のチェック日時 = 今回のチェック日時 + 最小更新間隔 / (連続未更新回数 + 1)」という計算を行うと、「11日」が得られ、この11日では見事に一致する。ここで、次回のチェック日時を検討する際には、「最小更新間隔 / (連続未更新回数 + 1)」というようにして徐々にチェック間隔を短くしている。これは、予定日を過ぎていてので、今後、更新される確率が高くなっていくことに鑑みて間隔を短くしていった。

【0090】次のチェック日時は「16日」であるが、14日に3日ぶりに更新されており、最小更新間隔を更新する。

を図8に示すシミュレーションを参照して説明する。

【0085】ここで情報取得効率 μ_i を以下のように定義し、本実施形態における更新予測を利用する場合と定期運転を利用する場合とを比較してみることにする。

【0086】

【0091】以後、自動設定では、「19日」、「21日」と空振りとなるものの、連続未更新回数のフィードバックが功を奏して「22日」には一致し、「28日」は一日遅れで情報を得られる。

【0092】一方、第一の定期パターンでは、「8日」において3日遅れで情報を取得し、「11日」に一致し、「15日」に1日遅れで情報を取得し、「18日」に空振りし、「22日」に一致し、「25日」に空振りする。

【0093】また、第二の定期パターンでは、「11日」に一致するものの、「5日」の更新情報を取得できなかった。そして、「18日」に4日遅れで情報を取得し、「25日」に3日遅れで情報を取得する。

【0094】以上のシミュレーションに基づいて有効性を検討した結果を図9に示している。アクセス回数に対する情報取得回数をヒット率と計算し、ペナルティの合計を加算しつつその平均を求めている。また、情報喪失数がある場合には、それも表示している。なお、参考のために毎日チェックした場合も含めてある。

【0095】さらに、図10には最終的な情報取得効率 μ_i を示しており、この更新予測を利用した場合の情報取得効率が最も高いことが分かる。なお、(1)式の特徴として、第二の定期パターンではヒット率が100%であるにもかかわらず、情報取得効率が低くなる傾向がある。これは、ペナルティが多かったのと情報の取りこぼしがあったことを重視しているからである。また、毎日チェックする場合にはペナルティが「0」となるもののアクセス数が非常に多いので効率が悪くなるという傾向も示している。いずれの場合も、実感的な効率に基づいて算出していることがよく分かる。

【0096】さて、これまではWEBサイトのHTML文書を取得する例について説明してきたが、このように情報を解析して更新予測する手法はこのようなHTML文書に限られるものではなく、図11に示すような自動的なシステムアップデート処理も適用可能である。

【0097】このシステムアップデート処理では、ファームウェアを含めた各種のシステムプログラムを自動的にダウンロードして更新させるものである。チェック日時の計算は前述した例に則っている。

【0098】ステップS600にて本日の日付と次回チェック日時として記憶されている日付とを比較し、一致していればステップS605にて情報を取得する。この情報の取得は、例えばファイルサーバに対して登録されているプログラムファイルの一覧表を取得する処理など

が該当する。そして、各プログラムファイルのタイムスタンプをチェックし、ステップS610にて更新されているものがあるか否かを判断する。更新されているものがあればステップS615にてそのプログラムファイルをダウンロード後、ステップS620にてシステムアップデート処理を実行する。このシステムアップデート処理は、ダウンロードするプログラムファイルを仮の領域に保存しつつ、システムをリセットし、イニシャル処理においてそのプログラムファイルを本来の領域にコピーしてから起動するといった手順などが考えられる。むしろ、同様の処理を別の手法によって実現することも可能であり、リセットしないで次の起動時にロードし直すようにしたり、リセットのタイミングを夜中などに行うようにするといったことも可能である。

【0099】更新したか否かは別のフラグなどに残しておき、ステップS625では更新の有無などに応じて次回チェック日時を設定する。

【0100】このように、従来であれば、勘に頼って定期的な自動運転を行ない、当てずっぽう式の手法で情報を取得するようにしていたが、得られた情報の中に次の更新予定日であるとか、更新間隔などがないか判断するとともに、これらが無い場合には過去の更新履歴に基づいて次の更新を予測するようにしたため、いたずらにアクセス回数を増やすことなく、可能な限り最新の情報を維持するということが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施形態にかかる情報取得装置のクレーム対応である。

【図2】本情報取得装置が広域ネットワークに接続され

る全体像を示す概略図である。

【図3】情報取得装置を実現するコンピュータ端末の概略構成図である。

【図4】情報取得装置を実現するソフトウェアの概略フローチャートである。

【図5】情報取得装置の設定用画面を示す図である。

【図6】初回アクセス処理のフローチャートである。

【図7】次会チェック日時取得処理のフローチャートである。

【図8】シミュレーションを示す図である。

【図9】シミュレーションの結果を示す図である。

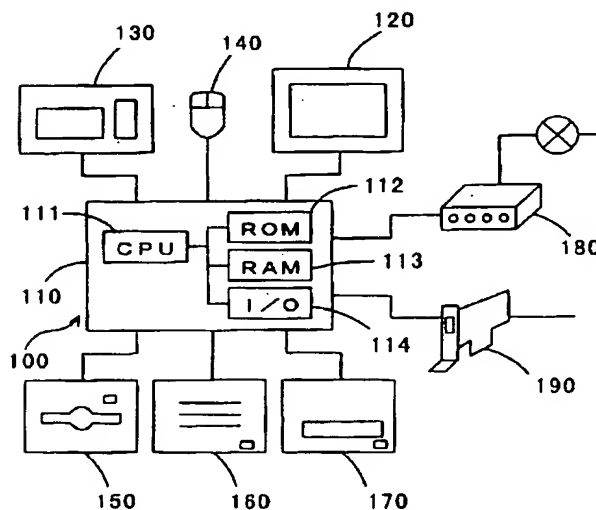
【図10】シミュレーションの結果に伴う情報取得効率を示す図である。

【図11】システムアップデート処理を示すフローチャートである。

【符号の説明】

- 10…情報取得装置
- 20…情報取得手段
- 30…自動運転手段
- 40…次回更新日時取得手段
- 41…更新予定日取得手段
- 42…更新間隔取得手段
- 43a…更新履歴記憶手段
- 43b…更新予測手段
- 43c…実更新日時フィードバック手段
- 50…広域ネットワーク
- 60…通信回線
- 70…情報発信源

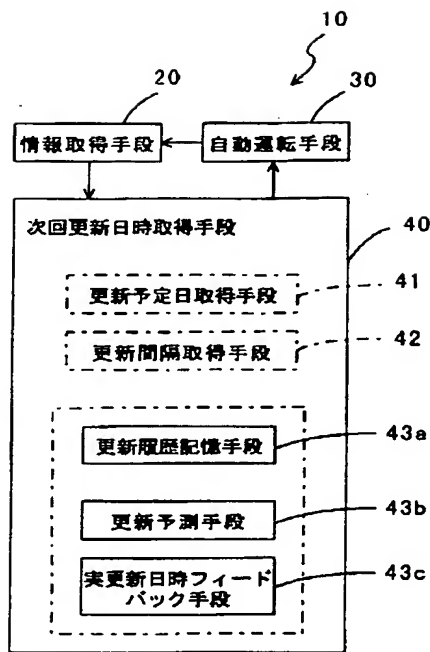
【図3】



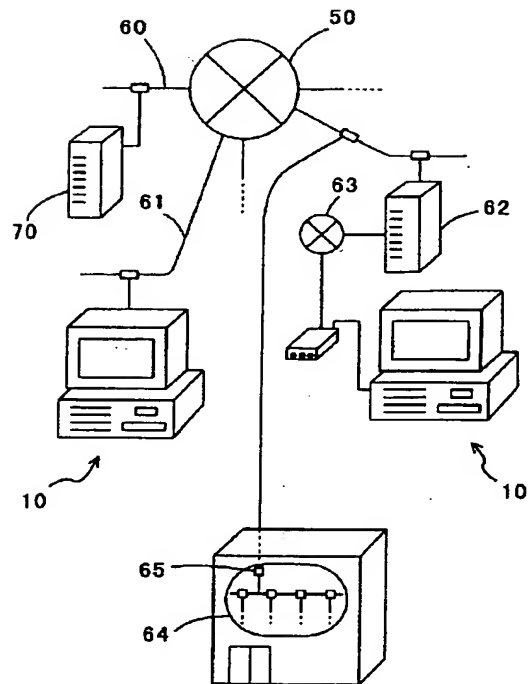
【図10】

巡回頻度	情報の取得回数	情報取得効率 (%)
更新予測	5	52.1
一日に一回	5	20.8
第一の定期パターン	4	42.3
第二の定期パターン	3	16.7

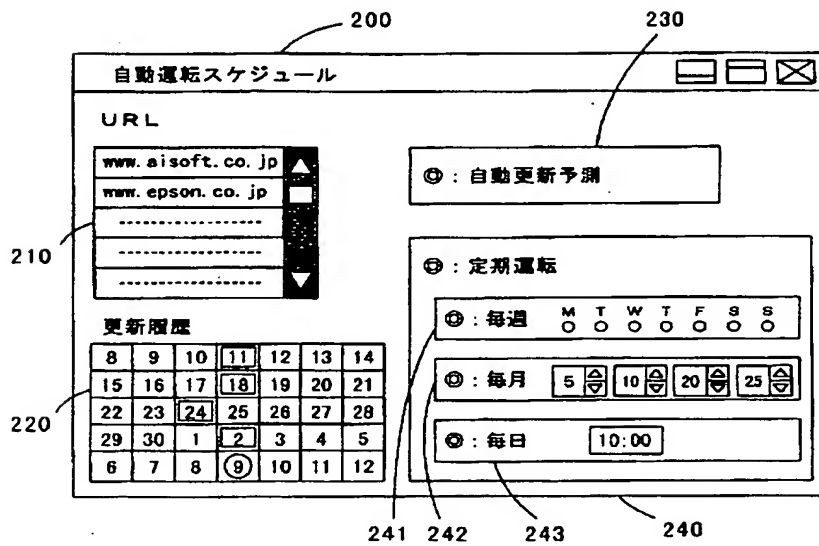
【図1】



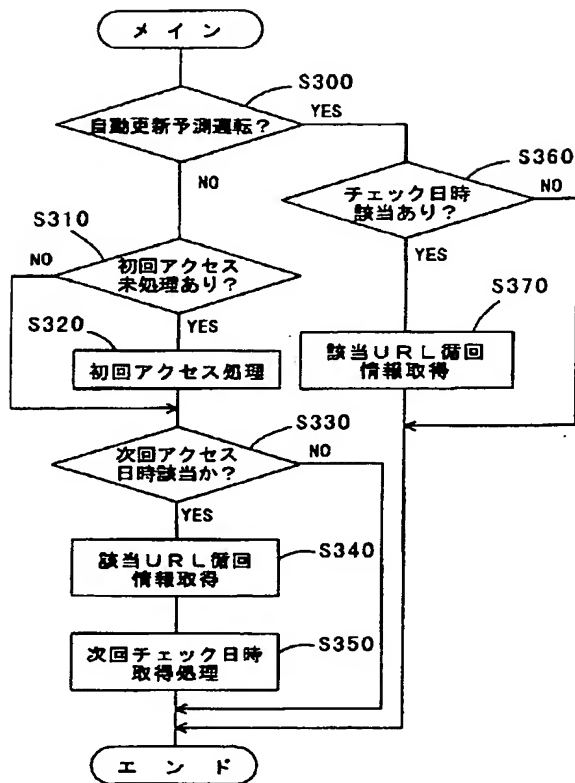
【図2】



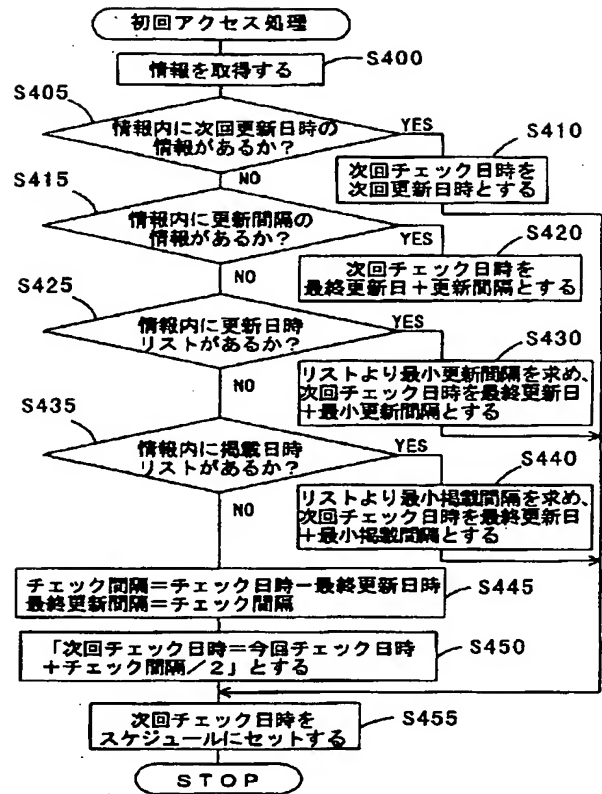
【図5】



【図4】



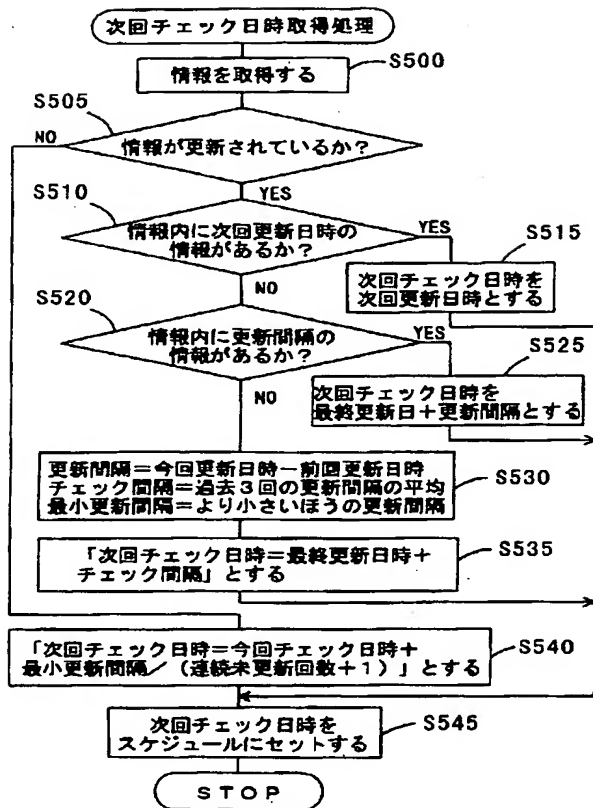
【図6】



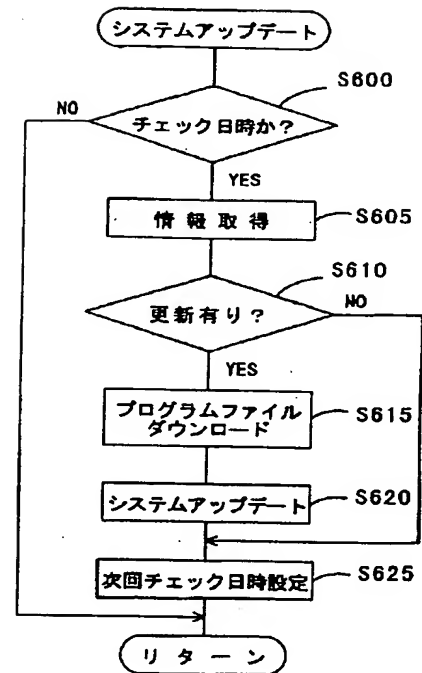
【図8】

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
更新日														
更新予測				*		⊙1			—		⊙0			
第一の定期パターン				*				⊙3			⊙0			
第二の定期パターン				*	x						⊙0			
	15	16	17	18	19	20	21	22	23	24	25	26	27	28
更新日														
更新予測		⊙2			—		—	⊙0						⊙1
第一の定期パターン	⊙1			—				⊙0			—			
第二の定期パターン				⊙4							⊙3			

【図 7】



【図 11】



【図 9】

巡回頻度	アクセス回数	情報取得回数	ヒット率(%)	ペナルティ	平均ペナルティ	情報喪失数
更新予測	8	5	62.5	4	0.8	0
一日に一回	24	5	20.8	0	0	0
第一の定期パターン	6	4	66.7	5	1.25	0
第二の定期パターン	3	3	100.0	8	2.33	1

フロントページの続き

(51)Int. Cl.⁶

識別記号

F I

G 0 6 F 15/74

3 2 0 D